

ADAM POUR LE DEEP LEARNING

ACHIQ Aya, CLETZ Laura, EL MAZZOUJI Wahel

Octobre 2025



UNIVERSITÉ DE
MONTPELLIER



STATISTIQUE
SCIENCE DES DONNÉES
UNIVERSITÉ DE MONTPELLIER



FACULTÉ DES SCIENCES
DE MONTPELLIER

SOMMAIRE

CADRE ET ENJEUX

FORCES ET FAIBLESSES

CONCLUSION

ENJEUX ET CADRE STATISTIQUE

- ▶ L'optimisation est **centrale** en Deep Learning : ajuster les poids θ pour minimiser une perte $\mathcal{L}(\theta)$.
- ▶ Méthodes classiques :
 - ▶ **SGD**, Momentum ;
 - ▶ Sensibles au **taux d'apprentissage** et aux gradients bruités.
- ▶ Enjeux :
 - ▶ Convergence rapide ;
 - ▶ Stabilité numérique ;
 - ▶ Bonne généralisation.
- ▶ Méthodes **adaptatives** : AdaGrad, RMSProp, puis **Adam** Kingma et Ba, 2014.

PRINCIPE DE L'ALGORITHME ADAM

IDÉE CLÉ

Adam = Adaptive Moment Estimation : combine le **momentum** et une adaptation du pas pour chaque paramètre.

- Moyennes mobiles :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

- Mise à jour :

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

- Valeurs typiques : $\alpha=0.001$, $\beta_1=0.9$, $\beta_2=0.999$.

Propriétés : invariance d'échelle, stabilité, faible sensibilité aux hyperparamètres.

OBJECTIFS ET LIMITES D'ADAM

Objectifs :

- ▶ Accélérer la convergence avec momentum et adaptation des pas ;
- ▶ Stabiliser l'apprentissage en présence de gradients bruités ;
- ▶ Adapter automatiquement le taux d'apprentissage pour chaque paramètre.

Limites :

- ▶ Moindre généralisation que SGD Wilson et al., 2017 ;
- ▶ Convergence vers $\|w\|_\infty$ minimale ;
- ▶ Compromis entre vitesse et généralisation.

EXPÉRIMENTATION

EXPÉRIMENTATION

CONCLUSION

RÉFÉRENCES

Kingma, Diederik P. et Jimmy Lei Ba (2014). « Adam : A Method for Stochastic Optimization ». In : url :

<https://arxiv.org/abs/1412.6980>.

Wilson, Ashia C. et al. (2017). « The Marginal Value of Adaptive Gradient Methods in Machine Learning ». In : url :

<https://arxiv.org/abs/1705.08292>.